MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 1, 2019/2020

DPS5018 –INTRODUCTION TO PROBABILITY AND STATISTICS

(For Diploma students only)

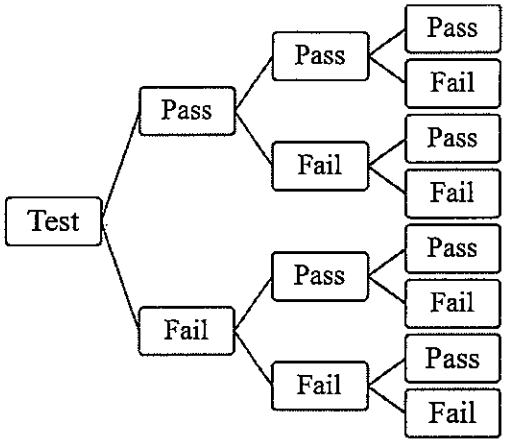
23 OCTOBER 2019
2.30 p.m. - 4.30 p.m.
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This question paper consists of **9** pages (**5** pages with **4** questions and **4** pages for appendix)
2. Answer **ALL** questions. All necessary working steps must be shown.
3. Write all your answers in the answer booklet provided.
4. Statistical Tables will be provided.

QUESTION 1 [15 Marks]

(a)



Possible Outcome: PPP, PPF, PFP, PFF, FPP, FPF, FFP, FFF

The possible outcome of three independent tests are given in the tree diagram above. Let X be the number of Pass result of the tests.

Outcome	FFF			PPP
X	0	1	2	3
$P(X=x)$	$\frac{1}{8}=0.125$			$\frac{1}{8}=0.125$

i. Reconstruct and fill in the probability distribution for X given above. (1 mark)

ii. From your observation, if Y be the number of Fail result of the tests, do you think that the probability distribution of Y will be the same as probability distribution of X? Justify your answer. (2.5 marks)

(b) Nowadays, the demand of software engineers is high in the IT industry. From a sample of eight high school students surveyed, 75% of them are very interested in IT field.

i. How many students are expected to be interested in IT field? (0.5 mark)

ii. Calculate the standard deviation of students that are interested in IT field. (1 mark)

iii. Find the probability that more than half of those surveyed are interested in IT field. (1.5 marks)

Continued.....

(c) On average, there are 7.3 defective microchips produced by factory A on a particular day. Help the production team to figure out the probability that factory A will produce

i. less than five defective microchips on a particular day. (1.5 marks)
ii. at most eight defective microchips on a particular day. (1 mark)
iii. more than seven defective microchips in two days. (2 marks)

(d) A squad team of PUBG gamers successfully qualified for the Global Championship 2019. The time spent of practicing on playing the PUBG game is normally distributed with a mean of twelve hours and standard deviation of five hours. From the previous champion's experience, the chance of winning the game is high if the gamers spend more time on practicing.

i. Find the probability that the gamers spend more than 24 hours playing PUBG game. (2 marks)
ii. Find the probability that the gamers spend around 10 to 21 hours on playing PUBG game. (2 marks)

QUESTION 2 [10 Marks]

(a) A random sample of 12 entrepreneurs at Tamarind Square were surveyed on the average cost per month on their shop's rental fee. The sample has a mean of RM1500 and a standard deviation of RM67 for the monthly expenses. Determine the 95% confidence interval for the corresponding population mean. (4 marks)

(b) Depression among Malaysians showed a significant increase in the number of depressed patients from 2011 to 2015. An expert wanted to test the current situation of depression in Malaysia. He needs to be 95% confident that the error is at most 0.05. How large his sample should be? (3 marks)

(c) Dyscalculia is a specific learning disability in mathematics. From a survey of 2000 residents in Selangor, 100 of them have heard of the disability.

i. Find the point estimate of the proportion. (1 mark)
ii. Calculate the margin of error of the estimation made for a 95% confidence level. (2 marks)

Continued.....

QUESTION 3 [15 Marks]

(a) In October last year, Malaysia Ministry of Transportation announced that child car seats would be mandatory for all private cars by 2020 to protect children and infants death due to road accidents. He believed that at least 45% of parents will support his plan. However, a study had been carried out this year and shows that the usage rate is still remains low at only 26% based on a sample of 80 private cars. *(Source: "Child seats to be made compulsory by 2020, says Loke"; Tuesday 30th October 2018; The Star Publication via www.thestaronline.com, "Usage rate of child car seats remains low"; Friday 26th April 2019; New Straits Time Publication via https://www.nst.com.my)*

i. State the null and alternative hypothesis. (1 mark)
ii. By using the hypothesis testing of one population proportion, calculate the test statistic. (1 mark)
iii. Is there any evidence to reject the claim? Test at 0.1% level of significance (2 marks)

(b) Exam Unit Head of Mol Mol University claims that students' performance has improved after implementing blended learning with an average CGPA of 3.14. A lecturer selected 20 students randomly and found that the sample mean is 3.23 with a standard deviation of 0.92. To validate this claim, test at 10% level of significance. (4.5 marks)

(c) Marketer believes that primary school children prefer to drink MILO as compared to secondary school teenagers. To validate the claim, two groups of students are being selected.

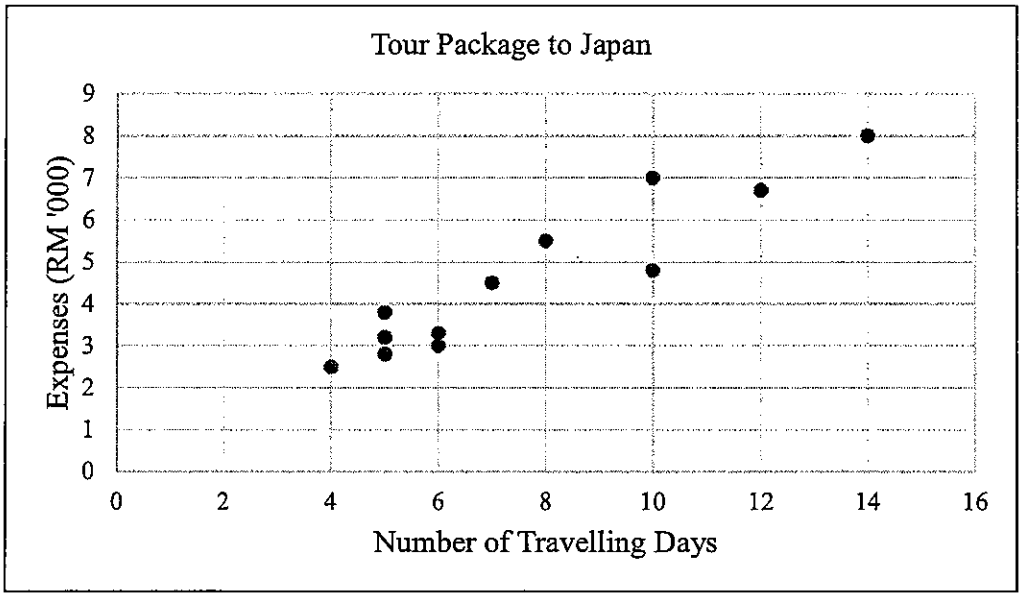
	Primary School	Secondary School
Number of students who prefer to drink MILO, x	43	32
Number of students being selected, n	50	50

i. Determine the samples proportion of primary school children, $\hat{p}_1$ and secondary school teenagers, $\hat{p}_2$ who prefer to drink MILO. (1 mark)
ii. State the null and alternative hypotheses. (1 mark)
iii. By using the hypothesis testing of two population proportion, compute the test statistic, Z. (2.5 marks)
iv. Test the claim at 5% level of significance. (2 marks)

Continued.....

QUESTION 4 [10 Marks]

(a) The graph below shows the number of travelling days and expenses of tour package to Japan for 12 persons. The expenses are subjected to accommodation, transportation, food or other expenses.



Number of Travelling Days, x	Expenses, y (RM '000)
7	4.5
6	3.0
10	4.8
4	2.5
k	7.0
14	8.0
5	3.2
5	3.8
8	5.5
6	3.3
5	2.8
12	6.7

Continued.....

i. Referring to the graph provided, a person spent RM7000 in k days. Determine the value of k. (0.5 mark)

ii. Calculate $\sum y$ and $\sum x^2$. (1.5 marks)

iii. Refer to part (a) ii., show that $SS_{yy} = 37.09$. (2 marks)

iv. Given that the $SS_{xx} = 110.67$ and $SS_{xy} = 60.27$, compute the coefficient of correlation, r. Then, comment your result. (2 marks)

v. How much a customer needs to prepare in order for him/her to travel in Japan for 20 days if the estimated least square regression equation is $\hat{y} = 0.38 + 0.55x$? (1.5 marks)

(b) Refer to part (a), the travel agency claims that there is a positive correlation between the number of travelling days and expenses of tour package.

i. State the null and alternative hypotheses. (1 mark)

ii. Given test statistic is 8.71, identify whether we should reject H_0 or fail to reject H_0 at 1% level of significance. (1.5 marks)

End of Page

APPENDIX – KEY FORMULA

- **Mean for ungrouped data**

$$\bar{x} = \frac{\sum x_i}{n} \quad \text{or} \quad \mu = \frac{\sum x_i}{N}$$

- **Mean for grouped data**

$$\bar{x} = \frac{\sum x_i f_i}{n} \quad \text{or} \quad \mu = \frac{\sum x_i f_i}{N}$$

Where x is the midpoint and f is the frequency of the class

- **Variance for ungrouped data**

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \text{or} \quad s^2 = \frac{\sum x_i^{\,2} - \frac{\left(\sum x_i\right)^2}{n}}{n-1}$$

- **Variance for grouped data**

$$s^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{n-1} \quad \text{or} \quad s^2 = \frac{\sum x_i^{\,2} f_i - \frac{\left(\sum x_i f_i\right)^2}{n}}{n-1}$$

Where x is the midpoint and f is the frequency of the class

DISCRETE RANDOM VARIABLE, BINOMIAL DISTRIBUTION, AND POISSON DISTRIBUTION

- Mean of a discrete random variable x : $\mu = \sum xP(X = x)$
- Standard deviation of a discrete random variable x : $\sigma = \sqrt{\sum x^2 P(X = x) - \mu^2}$
- Binomial probability formula : $P(X = x) = \binom{n}{x} p^x q^{n-x}$
- Poisson probability formula : $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$

- Key formula for Binomial Probability Distribution:
  - If $p \leq 0.5$
    i. $P(X = r) = B(r) - B(r-1)$
    ii. $P(X \geq r) = 1 - B(r-1)$
    iii. $P(X \leq r) = B(r)$
    iv. $P(X > r) = 1 - B(r)$
    v. $P(X < r) = B(r-1)$
    vi. $P(a \leq X \leq b) = B(b) - B(a-1)$

  - If $p > 0.5$
    i. $P(X = r) = B(n-r) - B(n-r-1)$
    ii. $P(X \geq r) = B(n-r)$
    iii. $P(X \leq r) = 1 - B(n-r-1)$
    iv. $P(a \leq X \leq b) = B(n-a) - B(n-b-1)$

- Key formula for Poisson Probability Distribution:
  - If $\mu = \lambda$
    i. $P(X = r) = Poi(r) - Poi(r-1)$
    ii. $P(X \geq r) = 1 - Poi(r-1)$
    iii. $P(X \leq r) = Poi(r)$
    iv. $P(X > r) = 1 - Poi(r)$
    v. $P(X < r) = Poi(r-1)$
    vi. $P(a \leq X \leq b) = Poi(b) - Poi(a-1)$

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value : $Z = \dfrac{x - \mu}{\sigma}$

ESTIMATION

- The $(1-\alpha)100\%$ confidence interval for μ is

$\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}$ where $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

$\bar{x} \pm t_{\alpha/2,n-1}s_{\bar{x}}$ where $s_{\bar{x}} = \dfrac{s}{\sqrt{n}}$

- The $(1-\alpha)100\%$ confidence interval for p is

$\hat{p} \pm z_{\alpha/2}s_{\hat{p}}$ where $s_{\hat{p}} = \sqrt{\dfrac{\hat{p}\hat{q}}{n}}$

- **Sample size**

$$n = \left[\frac{\left(Z_{\alpha/2}\right)\sigma}{E}\right]^2 \qquad n = \frac{\left(Z_{\alpha/2}\right)^2 pq}{E^2}$$

HYPOTHESIS TESTING

- **ONE population**

Mean:

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \quad \textbf{or} \quad t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

Proportion:

$$Z = \frac{\hat{p}-p}{\sqrt{\dfrac{pq}{n}}}$$

- **TWO population**

Mean:

$$Z = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$$

Proportion:

$$Z = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\hat{p}_p(1-\hat{p}_p)}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \qquad \text{where} \qquad \hat{p}_p = \frac{x_1+x_2}{n_1+n_2}$$

REGRESSION

- For the **least square regression equation**, $\hat{y} = b_0 + b_1 x$

  where $b_1 = \dfrac{SS_{xy}}{SS_{xx}}$ and $b_0 = \dfrac{1}{n}\left[\sum y - b_1 \sum x\right]$

- **Correlation coefficient**, r

  $$r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}}$$

  where

  $$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n}, \qquad SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} \text{ and}$$

  $$SS_{xy} = \sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n}$$

- **Regression *t*-Test**

  $$t = \frac{b_1 - \beta_1}{s_b} \qquad \text{where } s_b = \frac{s_e}{\sqrt{SS_{xx}}} \qquad \text{and } s_e = \sqrt{\frac{SS_{yy} - b_1 SS_{xy}}{n-2}}$$

- **Correlation *t*-Test**

  $$t = \frac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$$